Live Objects - Collaborative Window in the Corporate Documents

Riste Stojanov, Marjan Georgiev, Vladimir Zdraveski,
Milos Jovanovik, and Dimitar Trajanov

Faculty of Computer Science and Engineering, Ss. Cyril and Methodius in Skopje, Macedonia
{name.lastname}@finki.ukim.mk

Abstract. In this paper, a document collaboration platform for enterprise environments is presented. It incorporates the collaboration, security, auditing and reuse features into the document editor, as a tool that has well known interface to the end users. The platform enables template definition with annotation of the collaboration and protection units, called live objects. Through the live objects it allows multiple employees with different privileges to work on the same document, without having to violate the enterprise policies and processes. This work aims to increase the employees' productivity by providing a collaboration windows in corporate documents.

Keywords: enterprise, document, collaboration, security, auditing, reuse.

1 Introduction

In the digitalized and globalized environment of today, the amount and quality of knowledge makes the difference between the corporations. The enterprise knowledge is obtained from its employees and stored in various systems and formats [18]. In the process of knowledge management, teamwork and collaboration is required for task accomplishment[14][25], and the tools need to provide multiple employees with efficient collaboration, enabling knowledge distribution, synchronization and conflict management. Employees from different departments and various privileges can often work together[28], and their knowledge should be protected from unauthorized access and manipulation. The trust in the knowledge management systems is crucial[13], and the process of auditing is used to preserve it. Also, there is a need for managed knowledge reuse, with references of the reused parts for modification synchronization.

Corporate knowledge is represented in various formats, from structured information, stored in traditional databases[11], to unstructured information stored in documents.

The knowledge should be organized according to corporation policies, processes and standards, that are designed to increase the efficiency of their employees and to minimize the operation costs. In order to do so, tools that optimize the collaboration, security, auditing and reusing of the knowledge should be provided.

N. Bassiliades et al. (eds.), *New Trends in Database and Information Systems II*,
Advances in Intelligent Systems and Computing 312, DOI: 10.1007/978-3-319-10518-5_6

In most enterprises, all these aspects are achieved mainly through customized applications, leveraging the advantages of the database technologies [11]. These systems work with structured data and use the table records as fine-grained units for collaboration, protection, auditing and reuse. However, their main disadvantage is that they impose strict rules for data entry, and the employees may not be able to transform their knowledge to that format.

Beside the structured knowledge in database systems, the documents contain the major portion of the organizational information utilized in the corporate processes and used by different users and applications [20]. The documents are designed to store unstructured data in the form of natural language text and drawings, and provide a medium that can accept all available employee knowledge. Knowledge management in documents imposes greater conflict possibility when multiple employees collaborate on the same document. The document is the unit of management, which prevents collaboration of employees with different roles, makes it difficult to detect the exact change in the auditing process, and is too coarse for reuse.

In this paper, we present a platform that optimizes the collaboration, security, auditing and reuse of the document content, with a part of a document as a knowledge granularity unit.

The paper is organized as follows: in Section 2 the document collaboration approaches and applications are summarized, after which, in Section 3 the general overview of the implemented platform is presented. In Section 4, the implementation of the system is presented, and a comparison with the other solutions and methodologies is elaborated in Section 5. A conclusion about the presented work is given in the last section of the paper.

2 Related Work

Corporate knowledge can be managed in various systems and formats, including the Database Systems (DB)[11], Document Management Systems (DMS)[24], Content Management Systems (CMS)[22] and Online Document Collaboration Systems (ODCS)[27]. All these systems enable efficient collaboration, security, auditing and knowledge reuse on a different granularity units. The database and the Content Management systems work with records and content units correspondingly, which provide data structuring based on the corporate needs, and the employees should adopt their knowledge to the system requirements. According to [20], the major amount of the enterprise knowledge is stored in documents. Even though the DBS and CMS can produce documents, they are not intended to manage them.

Currently, there are two general document collaboration patterns. The first pattern is represented by the online document collaboration systems[27], where the employees edit the document in that system. The second pattern is when the documents are protected and distributed by one system, and edited by another, represented by document management systems.

The online document collaboration systems[27], such as Google Drive[1] and Microsoft Office 365[17], provide real-time collaboration among the employees

and are deployed in public cloud infrastructures. These systems provide similar set of features as most of the modern document editors, and there is simple or no training required for employees to use them[27]. The protection granularity of these systems is the document, and there is no mechanism to protect part of content from employees that do not have the rights on it. Additionally, their infrastructure may violates the security policies of many corporations that do not allow storing their documents outside their infrastructure. Even if they do allow this, the corporate accounts should be connected with the platform accounts, which is not that simple process.

The Please Review[3] is another online document collaboration system that can be installed in the corporation. The primary goal of this system is collaborative review, but it also offers a co-authoring option. In the process of document collaboration, the document owner shares the document with the collaborators and assigns them in a so called editing zone. These zones are locked until the user is done with his work. The zone should be downloaded, edited and uploaded back to the system, since it does not provide web editing. Additionally, the user can see the whole document, and there is no option to protect only a part of the document with this system.

The DMS, that represent the second pattern are more often used in the corporations. The collaboration unit of these systems is the document. The document management systems, such as SharePoint[21], Alfresco[23] and many other[19][8][15][16], provide collaboration by document distribution and versioning. Conflicts are prevented by document locking, which disables concurrent modification and decreases the efficiency. Such systems provide document level security and auditing with corporate account integration. These systems does not manipulate with the content of the document, and are combined with document editing applications for content manipulation.

Similar approach is utilized in the file synchronization systems such as Dropbox[9], OneDrive[2] and Ubuntu One[4]. These systems provide document level security and client applications for document synchronization. However, all these solutions are cloud based and may be against the corporate rules.

In [6][5][10] document collaboration systems and methods are defined, where the document is split in workable elements that can be edited by multiple collaborators and the changes are synchronized with a database system in real-time using messaging system. These systems use document level protection using single passwords, and perhaps niter of them provide annotation of the workable elements according to their permissions or meaning in a document template. Also, they do not provide any tracking of the reused parts and are web based solutions.

3 Semantic Sky Platform

In corporate environments, the documents frequently contain parts that should not be accessible to some employees. Since none of the systems reviewed in Section 2 protect part of a document, they enforce document splitting based

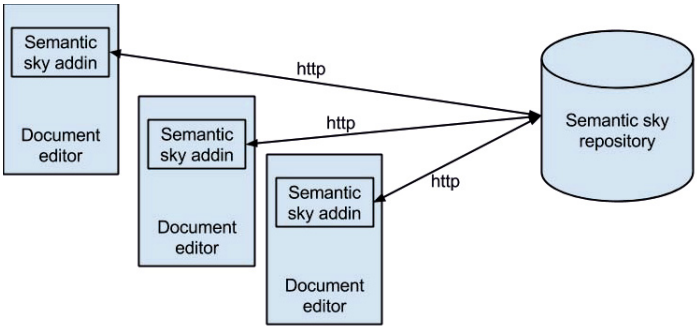

Fig. 1. Semantic Sky platform architecture

on the user privileges, collaborating on each of them and merging back the modifications.

The platform presented in this paper, called Semantic Sky, extends our previous work [26], by providing collaboration, security, auditing and managed reuse of document parts. The document parts managed by the Semantic Sky platform are referred to as live objects.

3.1 Semantic Sky Architecture

The Semantic Sky platform is implemented using client server architecture (Figure 1). The clients are implemented as document editor addins, while the server is RESTful live object repository. The platform is designed to coexist with the document management systems[24], and the current implementation of the addin is integrated with SharePoint[21] DMS, but not bounded to it.

The Semantic Sky addin is used to handle the document templates and live objects, and currently is implemented for Microsoft Office Word, as one of the most common document editing applications in the corporate world. Moreover, almost all corporate employees are trained to use document editors and they can use the addin with short or no additional training. This way, the employees can use the tools they know best, and simultaneously modify the document parts required to accomplish their tasks. The Semantic Sky addin is used to detect and manipulate the document's live objects, synchronize their content among the collaborators, and protect them from unauthorized access and modification.

All Semantic Sky client addins communicate with the central repository to provide synchronization of the live object content among the users. The content of the live object is versioned, and multiple versions for each live object are persisted into the repository. The Semantic Sky addin displays the content from the last version of the live object in the document. The live object versions provide a better insight to the editors about the evolution of the document. The users have an option to preview the versions of a selected live object, and a possibility to select the content of some previous version as current.

3.2 Live Objects

The enterprises use documents with standardized structure in a form of document templates. The templates define the structure of the document, where each part has position and meaning. The Semantic Sky platform is designed to coexist with the templates, where the dynamic and protected parts can be annotated as live objects.

The live objects wrap parts of a document that represents a logical unit and allow collaborative editing. The best practice is to wrap self-contained logical units for collaboration, such as paragraphs, sentences or chapters in live object, but the platform does not impose any limitation.

The live objects are the main units for collaboration and they represent the meaning of the element in the template, while their versions store the content, the modifier and the modification time. The Semantic Sky system uses an adaptation of the centralized version control[7] collaboration model for synchronization and conflict resolution with the live objects as granularity unit. The system automatically invokes the update and commit procedures on document open and save correspondingly.

The live objects are designed to enable managed reuse, where the reusing live object references to the reused one. This provides another level of document linking, and gives information about their content sharing. Additionally, when the reused live object is modified in some document, the documents that have reused it will obtain a notification for the modification, with option to accept it or not.

In order to provide collaboration of employees with different privileges in a same document, the Semantic Sky platform has separate access permissions for each live object. The system uses the RBAC model[12], where the users and their roles are obtained from theirs' computer account. Since the documents can be opened without the Semantic Sky addin, there are both public and private protection modes for the live objects. The public live objects are visible to everybody and their content is embedded into the document, while the content of the private live objects is synchronized on each document opening. For the private protection mode, the Semantic Sky system defines the following privileges: no access, read only, read & write, and manage. The higher privileges include and extends the rights from the lower one. The user that creates the document has modify permission, and can manage the protection mode and the privileges of the other collaborators. Live objects have public protection mode by default, and every user can read and modify their content.

3.3 Semantic Sky Document Editor Addin

The user interface of MS Word instance with the Semantic Sky addin is shown in Figure 2. All live objects are accessible through the panel at the left side of the editor, providing easy navigation to the required parts. There is also an option to preview the live objects from the other documents and to reuse their content in the current document (the right tab of the panel). The addin provides managed

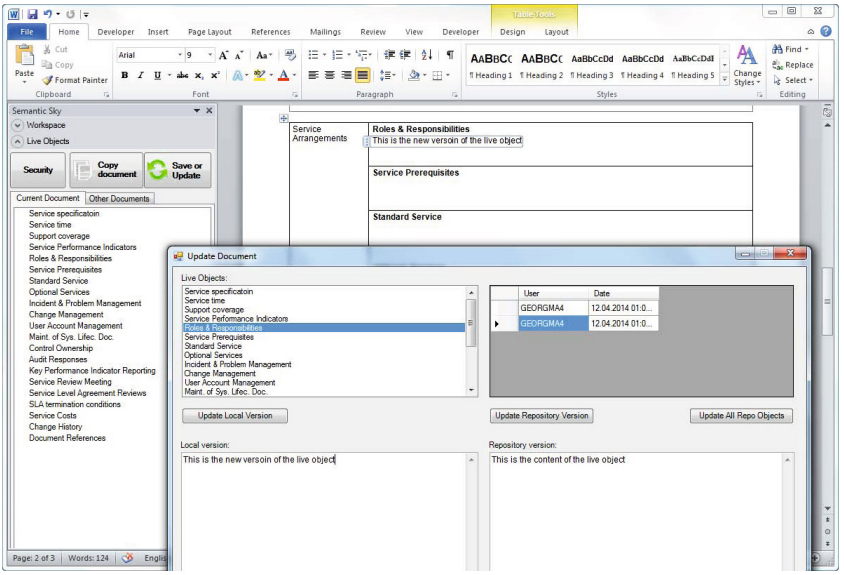

Fig. 2. Document with live objects and their versions

reuse, where the reusing live object stores a link to the reused one. The auditing and the conflict resolution is done through the popup window from Figure 2. It displays a list of all live object changes and enables to roll back the modifications to a previous version. The Semantic Sky platform provides access to the content even if the addin is absent, so that external collaborators can be included. The modifications from these collaborators are synchronized on the next document opening in document editor with installed Semantic Sky addin.

The platform does not provide real-time synchronization because the document may be opened in parallel by document editor instances (or other applications) that do not have the addin. The addin provides on demand live object synchronization and mechanism for conflict resolution. This way, the platform ensures the consistency of the live objects content among the different collaborators.

4 Semantic Sky Implementation

The live objects are implemented as MS Word native Rich Text Content Controls[1], handled by the semantic sky addin. The addin supports template creation through annotation of the live objects that are used for content protection and collaboration. The semantic sky addin uses the Tag property[2] of the Rich Text

[1] http://msdn.microsoft.com/en-us/library/microsoft.office.tools.word.richtextcontentcontrol.aspx

[2] http://msdn.microsoft.com/en-us/library/microsoft.office.tools.word.richtextcontentcontrol.tag.aspx

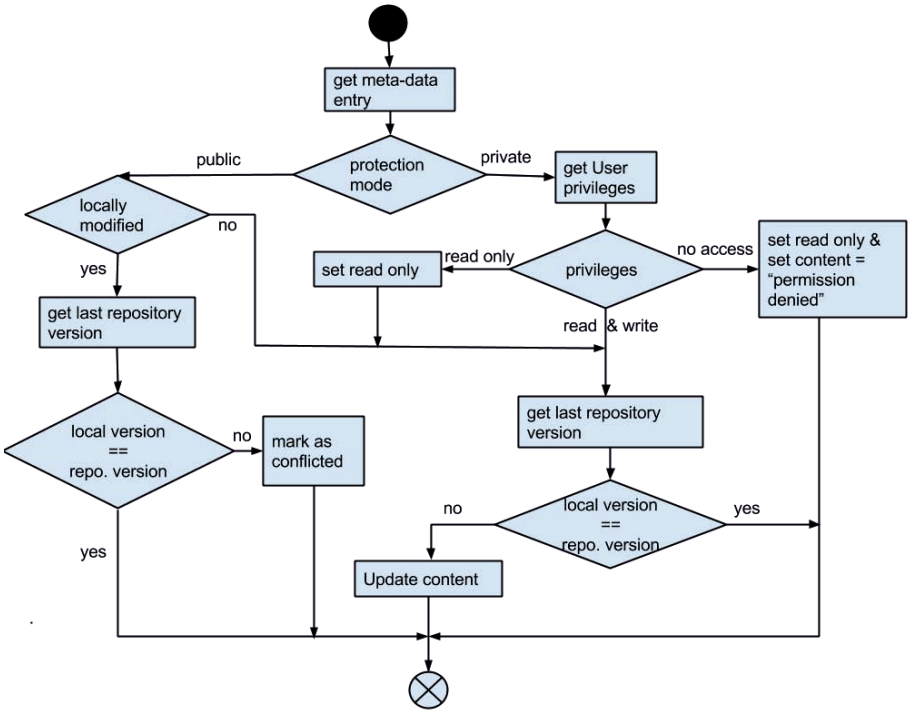

Fig. 3. Live Objects Initialization

Content Control to identify the live object controls, and the CustomXMLPart[3] to store their meta-data. The CustomXMLPart is defined as part of the Office Open XML standard[4], and it is not visible to the collaborators. The Semantic Sky addin uses it to decide how to protect and handle the live objects.

4.1 Live Object Initialization

When the document is opened in MS Word with Semantic Sky addin, the live objects mechanism iterates over all Reach Text Content Controls in the document, identifying the one that has Tag property starting with "#ssky|#lo|". For each live object control, the addin executes the initialization procedure from Figure 3 in a separate thread, since there are no correlations between them. This speeds up the document opening process, and each live object is initialized independently.

During the initialization process, the addin first obtains the live object's meta-data entity from the document's CustomXMLPart. Since the live objects with public protection mode can be modified outside of the system, there may be a

[3] http://msdn.microsoft.com/en-us/library/bb608618.aspx
[4] http://www.ecma-international.org/publications/standards/Ecma-376.htm

conflict in their content, and the addin checks whether they are modified in the document and in the repository. When there is conflict for the live object, it is marked as conflicted by setting this flag in its meta-data entry. The public live objects that are not locally modified, are update their content with the repository version.

The modification of the private live object is managed by the addin and there is no need for conflict resolution in their initialization phase. The access permissions for the user are first obtained for them. When the user does not have access for the live object, the read only property for its control is set and its content is not loaded nor displayed. Otherwise, the content is loaded from the repository only if the repository version is greater than the local version of the live object.

At the end of the initialization process, the addin registers a handler for the Exiting Event[5] of the live object control, which updates the live object's meta-data after each user interaction.

4.2 Live Object Synchronization

The content of the live objects can be synchronized on demand and on document save operations. In this process, the addin tries to store only the modified (dirty) live objects and tries to store their content in the repository. The system then checks for conflicts, and when there is none, a new version of the live object is created. Then the addin updates the live object's content and meta-data. In case the local version is smaller than the repository version, the repository stores the user live object version as conflicted, and the addin also marks the live object control as conflicted.

The popup dialog from Figure 2 is used for conflict resolution. On the upper left side all live objects are displayed, with the conflicted one colored red. In the lower left text box, the local version of the live object is displayed and the right text box is the latest repository version from the live object. The user can accept either the local or the repository version. If a combination from the both versions is needed, the user should manually change the content of its local version, and then accept it. The repository content (the right text area) is not editable. This dialog should be improved in order to display where the two versions differ, and it is planned to be improved as a part of the future work.

When the document is closed and all live objects are synchronized, the semantic sky addin removes the content from the live object controls with private protection mode. This way, when the document is opened without the addin, the protected live objects are not displayed and their information is protected.

5 Evaluation

The main difference of the Semantic Sky system is that it supports managed content reuse with referencing and protects the document parts from

[5] `http://msdn.microsoft.com/en-us/library/microsoft.office.tools.word.contentcontrolbase.exiting.aspx`

unauthorized access. All other systems provide content reuse with copy-paste, and do not provide tracking of the reused parts. Also, none of these systems provides protection of document parts based on employee privileges, and these employees must use workarounds that decrease their productivity in order to collaborate with these systems.

The on demand synchronization of the Semantic Sky system is not as efficient as the real-time synchronization, but it allows external collaborators to contribute in document editing tasks on a same document, without exposing protected information. All other systems either protect the whole document with password or do not provide protection at all.

Additionally, the Semantic Sky platform is designed to coexist with the DMS[24] and to use them for document distribution among the employees. The collaboration environment is integrated in Microsoft Office Word document editor, and the employees can use the platform with minimal training, and increases their efficiency.

Table 1 summarizes the features of the document collaboration systems. The collaboration feature describes the collaboration unit of each of the systems together with the synchronization mechanism. The security and auditing columns display the protection and versioning unit of each of the systems, while the content reuse column describes how content can be reused in them. The last column describes whether and how the documents are protected when they are being open out of the system.

Table 1. Document collaboration systems comparison

	Collaboration	Security	Auditing	Content reuse	Template support	Out of system protection
Semantic Sky	live object, on demand	live object	live object	referencing (managed)	yes	private live objects
ODCS[27]	character, real-time	document	character	copy-paste	no	no
DMS[24] with editing app.	document, lock&versions	document	document	copy-paste	editor templates	document (password)
[3][5][6][10]	document part, lock or real-time	document	document part	copy-paste	no	document (password)

6 Conclusion

The Semantic Sky platform provides document collaboration for enterprise environments, handling security with document content protection, even when the documents leave the controlled environment. The template mechanism enables structuring of the corporate documents through annotation of the collaborative and protected elements, called live objects.

The platform extends the document editors, with the Semantic Sky addin. The integration with the document editor significantly improves the employees efficiency because they can focus on their tasks, leaving the addin to handle the document synchronization and management. The versions of the live objects provide document evolution auditing, while the managed live object reuse provides notifications for the documents with reused content when the base live object is changed.

The use of the live object controls enables collaborators without the semantic sky addin to be involved in the editing process of the public content, without violating the corporate policies.

References

1. Google drive, `http://drive.google.com/` (accessed April 10, 2014)
2. Onedrive, `https://onedrive.live.com/about/en-us/` (accessed April 10, 2014)
3. Please review, `http://www.pleasetech.com/pleasereview.aspx/` (accessed April 10, 2014)
4. Ubuntu one, `https://one.ubuntu.com/` (accessed April 10, 2014)
5. Bailor, J., Bernstein, E., Knight, M., Antos, C., Simonds, A., Jones, B., Clarke, S., Sunderland, E., Robins, D., Bose, M.: Collaborative authoring. US Patent 7,941,399 (May 10, 2011)
6. Chin, R., Lee, J.: Document collaboration system and method. US Patent App. 11/836,087 (March 6, 2008)
7. Collins-Sussman, B., Fitzpatrick, B., Pilato, M.: Version control with subversion. O'Reilly Media, Inc. (2004)
8. Cullen, J., Peairs, M.: Document management system. US Patent 5,893,908 (April 13, 1999)
9. Drago, I., Mellia, M., Munafo, M.M., Sperotto, A., Sadre, R., Pras, A.: Inside dropbox: understanding personal cloud storage services. In: Proceedings of the 2012 ACM Conference on Internet Measurement Conference, pp. 481–494. ACM (2012)
10. Dutta, K.: Document collaboration system. US Patent App. 10/900,807 (February 2, 2006)
11. Elmasri, R.: Fundamentals of database systems. Pearson Education India (2008)
12. Ferraiolo, D., Cugini, J., Kuhn, D.R.: Role-based access control (RBAC): Features and motivations. In: Proceedings of 11th Annual Computer Security Application Conference, pp. 241–248 (1995)
13. Ford, D.P.: Trust and knowledge management: the seeds of success. In: Handbook on Knowledge Management 1, pp. 553–575. Springer (2004)
14. Goh, S.C.: Managing effective knowledge transfer: an integrative framework and some practice implications. Journal of Knowledge Management 6(1), 23–30 (2002)
15. Hajmiragha, M. Document management system. US Patent 6,289,460 (September 11, 2001)
16. Jeffery, S., O'Gwen, G., Hornsby, B., McBryde, K., Powell, W., Rizk, T.: Document management system. US Patent 6,957,384 (October 18, 2005)
17. Katzer, M., Crawford, D.: Office 365: Moving to the cloud. In: Office 365, pp. 1–23. Springer (2013)
18. O'Leary, D.E.: Enterprise knowledge management. Computer 31(3), 54–61 (1998)

19. Oliszewski, M.: Document management system. US Patent App. 10/208,062 (July 3, 2003)
20. Päivärinta, T., Tyrväinen, P.: Documents in information management: Diverging connotations of'a document'in digital era. In: Proceedings of IRMA 1998, pp. 163–173 (1998)
21. Pattison, T., Connell, A., Hillier, S., Mann, D.: Inside Microsoft SharePoint 2010. Microsoft Press (2011)
22. Rockley, A., Kostur, P., Manning, S.: Managing enterprise content: A unified content strategy. New Riders (2003)
23. Shariff, M.: Alfresco enterprise content management implementation. Packt Publishing Ltd. (2007)
24. Sutton, M.J.D.: Document Management for the Enterprise: Principles, Techniques, and Applications. John Wiley & Sons, Inc., New York (1996)
25. Syed-Ikhsan, S.O.S., Rowland, F.: Knowledge management in a public organization: a study on the relationship between organizational elements and the performance of knowledge transfer. Journal of Knowledge Management 8(2), 95–111 (2004)
26. Trajanov, D., Stojanov, R., Jovanovik, M., Zdraveski, V., Ristoski, P., Georgiev, M., Filiposka, S.: Semantic sky: a platform for cloud service integration based on semantic web technologies. In: Proceedings of the 8th International Conference on Semantic Systems, pp. 109–116. ACM (2012)
27. Vallance, M., Towndrow, P.A., Wiz, C.: Conditions for successful online document collaboration. TechTrends 54(1), 20–24 (2010)
28. Wong, K.Y., Aspinwall, E.: Development of a knowledge management initiative and system: A case study. Expert Systems with Applications 30(4), 633–641 (2006)